GMR The Original

Where Communities Preserve Genes: SCoT-Based Genetic Assessment of Dipterocarpus alatus Roxb. Ex G.Don in Thailand

Preeya P. Wangsomnuk1, Xin Chun Mo[2], Manop Poopath[3], Piyorot Honsachart[4]

[1]Department of Biology, Faculty of Science, Khon Kaen University, Khon Kaen, Thailand 40002. prepua@kku.ac.th

[2]Yunnan Economics Trade and Foreign Affairs College, Kunming, China 651700. ljnumxc@163.com

[3]Forest Herbarium, Department of National Parks, Wildlife and Plant Conservation, Chatuchak, Bangkok, Thailand 10900. poomanop@hotmail.com

4Department of Applied Sciences, Faculty of Interdisciplinary Studies, Khon Kaen University, Nong Khai Campus, Nong Khai, Thailand 43000 piyorot@kku.ac.th

Corresponding author: P.P. Wangsomnuk. E-mail: prepua@kku.ac.th

ABSTRACT

Dipterocarpus alatus Roxb. ex G.Don is an ecologically and economically important tree species in Thailand, but its natural populations are increasingly affected by habitat loss and fragmentation. Understanding its genetic diversity under these pressures is essential for effective conservation planning. Using start codon targeted (SCoT) markers, we assessed genetic variation in 219 individuals sampled from 14 district-level populations that represent a broad span of the species' distribution. Seven primers produced polymorphic profiles with an average PIC of 0.41. Population differentiation was moderate (PhiPT = 0.156), and AMOVA showed that most genetic variation (84%) occurred within populations. A significant but moderate isolation-by-distance pattern was also detected, indicating that geographic distance contributes to spatial genetic structure in D. alatus. STRUCTURE analysis resolved five main genetic clusters, with some individuals showing signs of admixture. While certain populations exhibited strong genetic homogeneity, others displayed evidence of recent gene flow. A notable portion of the sampled individuals originated from community forests and village-managed areas, highlighting the important—yet often under-recognized—role of these local landscapes in maintaining genetic resources. Although SCoT markers have inherent limitations, they proved useful for revealing meaningful genetic patterns in this non-model species. Overall, the results provide a sound basis for conservation strategies that link formal protected areas with locally managed forests.

Keywords: Dipterocarpus alatus, SCoT markers, genetic diversity, population structure, community forest, conservation genetics

INTRODUCTION

Dipterocarpus alatus Roxb. ex G.Don, belonging to the Dipterocarpaceae family, is found throughout tropical Asia, embracing India and mainland Southeast Asia; it is a big tree and essential to support the local economies and environmental sustainability; consequently, it is a vital species for this region (Orwa et al., 2009; Ly et al., 2023). The species is frequently sighted in Thailand's lowland deciduous and riparian forests, especially in the north, northeast, and central zones. *Dipterocarpus alatus* provides vital support that maintains forest structure, watershed steadiness, and local community livelihoods, historically used for timber, resins, and herbal medicines (Dyrmose et al., 2017; Kamyo and Asanok, 2019).

Notwithstanding its environmental and local community financial significance, the population of *D. alatus* has declined markedly in recent times due to habitat fragmentation, alterations in land use, and ecological overshoot. The IUCN Red List (Ly et al., 2023) currently classifies it as "Near Threatened". Several Thai studies raise concerns about the population's genetic integrity. Chokthaweepanich et al. (2022) assessed the *D. alatus* populations in Sakon Nakhon province using SRAP and SSR markers; the study revealed low observed heterozygosity (Ho = 0.163) and consistently positive inbreeding coefficients. However, the study had limitations, as the geographic extent was constrained to 87 trees within a 20 km area, thereby limiting the universality of the findings. A nationwide isozyme study conducted by Changtragoon and Boontawee (1999), which analyzed seeds from 50 trees across 16 natural populations, similarly reported low levels of heterozygosity (He = 0.092–0.094) and moderate genetic differentiation. Although informative, isozyme markers are still inherently limited in their resolution and unable to detect subtle genetic variations due to the relatively small number of loci involved.

In recognition of these genetic concerns, *D. alatus* has been designated as one of five priority species for urgent conservation under Thailand's Master Plan for Forest Genetic Resources Management (Changtragoon et al., 2013). While in situ conservation is considered ideal, many natural populations exist outside protected areas. Consequently, ex-situ strategies, including the establishment of genetic conservation plots, have been adopted to preserve genetic resources for future restoration (Thianchai et al., 2013).

Interestingly, studies conducted in Vietnam utilizing RAPD and SSR markers have revealed moderate to high levels of genetic diversity, along with considerable within-population variation in *D. alatus* (Nguyen et al., 2012; Tam et al., 2014; Vu et al., 2019). These interregional differences may reflect variations in ecological conditions, landscape histories, or conservation practices. Collectively, these findings underscore the importance of re-evaluating the genetic structure of *D. alatus* in Thailand through broader geographic sampling and more informative molecular marker systems.

SCoT markers, developed by Collard and Mackill (2009), are gene-targeted, highly reproducible, and cost-effective molecular tools that enable the assessment of functional genetic variation without the need for prior sequence information. These markers amplify regions adjacent to the ATG start codon, thereby preferentially targeting coding regions, which makes them particularly valuable for genetic studies in non-model species. Since their introduction, SCoT markers have been widely utilized across numerous plant species for various applications, including the assessment of genetic diversity, analysis of population structure, and investigation of interspecific or intergeneric relationships. In tree species, successful applications have been reported in *Tecomella undulata*, *Prosopis cineraria*, *Pinus bungeana*, and *Acer monspessulanum* (Rai, 2023). Despite their widespread application, there are currently no published studies reporting the use of SCoT markers in any species within the Dipterocarpaceae family. To the best of our knowledge, the present study represents the first investigation into the application of this marker system in this ecologically and economically significant group of tropical trees.

To overcome the current limitations in marker resolution and spatial representation, this study utilized SCoT markers to analyze the genetic diversity and population structure of *D. alatus* across 14 districts spanning northeastern, central, and western regions of Thailand. Field sampling integrated both phenotypic identification and local ecological knowledge, and included a formally designated genetic conservation site. This study specifically aimed to (1) evaluate genetic diversity using SCoT markers, (2) assess population structure and genetic relationships among individuals from different geographic regions, and (3) investigate spatial-genetic patterns to inform both in situ and ex situ conservation strategies.

MATERIALS AND METHODS

Plant Materials and DNA Extraction

Field surveys and interviews with local communities were conducted to identify naturally occurring *D. alatus* trees based on observable phenotypic traits, such as tree morphology, leaf shape, floral features, and fruit characteristics, supported by indigenous knowledge (Wati et al., 2024). Sampling was conducted in accessible community-based locations, including agricultural fields, temples, schools, residential areas, public spaces, and government-owned properties (Caneva et al., 2020). Furthermore, trees from the *D. alatus* Genetic Conservation Area within Khon Kaen University (designated as population KK2) were included to represent locally managed conservation resources. All sampling sites were located outside designated protected forest areas and therefore did not require official collection permits. When needed, written permission was obtained from landowners or local authorities. All field activities complied with Thai national laws and institutional guidelines.

From November 2017 to May 2023, samples were collected from 219 mature individuals across 14 populations, each corresponding to a district-level location within 10 provinces of Thailand (Table 1, Figure 1A–B). The sampling sites were distributed across three geographical regions: northeastern Thailand (Nong Khai, Udon Thani, Nakhon Phanom, Sakon Nakhon, Khon Kaen, Roi Et, and Sisaket), central Thailand (Nakhon Sawan and Sing Buri), and western Thailand (Phetchaburi). The sampling spanned 14 districts and 33 subdistricts, with site elevations ranging from 17.9 to 225.8 m above sea level.

Multiple populations represented provinces such as Roi Et and Khon Kaen to capture geographic and ecological heterogeneity. The number of populations per province reflected both the species' natural distribution and the accessibility of *D. alatus* individuals in the field. Geographic coordinates (latitude and longitude) and elevation were systematically recorded at each sampling site. Each population contained multiple sampling points within the same district. The minimum and maximum values of the coordinates and elevation are summarized in Table 1 to show the spatial range of each district. For Figure 1, a single representative location per population was mapped using the mean latitude and longitude of all sampling points within that district.

All sampled individuals had undetermined genetic backgrounds and were selected in collaboration with local communities and conservation managers to represent natural and ecologically meaningful diversity. Leaf or bark tissues were collected for DNA extraction and downstream molecular analyses.

Table 1. Population information of *D. alatus* from 14 populations in Thailand.

Pop No.	Pop Code	Province	District	Subdistrict (n)	No. of Samples	Latitude (°N)	Longitude (°E)	Elevation (m)
1	NK	Nong Khai	Pho Tak	Pho Tak (7)	7	17.85172 – 17.85217	102.45831 – 102.45862	175 – 176
2	UT	Udon Thani	Phen	Sum Sao (4)	4	17.58072 – 17.58092	103.01375 – 103.01389	184 – 185
3	NP	Nakhon Phanom	Mueang Nakhon Phanom	Kham Thao (19) Tha Kho (4) Nong Yat (5)	28	17.25337 – 17.39132	104.75380 – 104.79592	146 – 170
4	SN	Sakon Nakhon	Nikhom Nam Un	Nikhom Nam Un (4) Suwannakham (24)	31	17.14679 – 17.19195	103.68563 – 103.73834	188 – 264
5	KK1	Khon Kaen		Nong Pling (3) Khao Noi (1)	13			182

			Wiang Kao	Mueang Kao Phatthana (5) Nai Mueang (7)		16.45889 – 16.70028	102.28751 – 102.81556	– 220
6	KK2	Khon Kaen	Mueang Khon Kaen	Nai Mueang (30)	30	16.44611 – 16.45961	102.815001 – 102.81613	162 – 187
7	RE1	Roi Et	Selaphum	Klang (13) Na Mueang (7)	20	16.02611 – 16.06671	103.93500 – 103.94444	134 – 142
8	RE2	Roi Et	Thawat Buri	Niwet (7)	7	16.03611 – 16.03750	103.74583 – 103.74639	146 – 150
9	RE3	Roi Et	Mueang Roi Et	Non Tan (6)	6	15.96528 – 15.96611	103.68028 – 103.68111	152 – 157
10	RE4	Roi Et	Mueang Suang	Nong Phue (14)	14	15.82214 – 15.82417	103.71707 – 103.71821	152 – 161
11	NS	Nakhon Sawan	Mueang Nakhon Sawan	Kriangkrai (7)	7	15.73002 – 15.73102	100.17954 – 100.18005	32 – 35
12	SB	Sing Buri	In Buri	Huai Chan (6)	6	14.99543 – 14.99634	100.27687 – 100.27750	16 – 39
13	Ssk	Sisaket	Benjalak	Siao (19)	19	14.77695 – 14.77780	104.66530 – 104.66639	141 – 147
14	Pcb	Phetchaburi	Ban Lat	Rai Makham (12) Tham Rong (15)	27	13.01174 – 13.01853	99.91420 – 99.92086	14 – 21

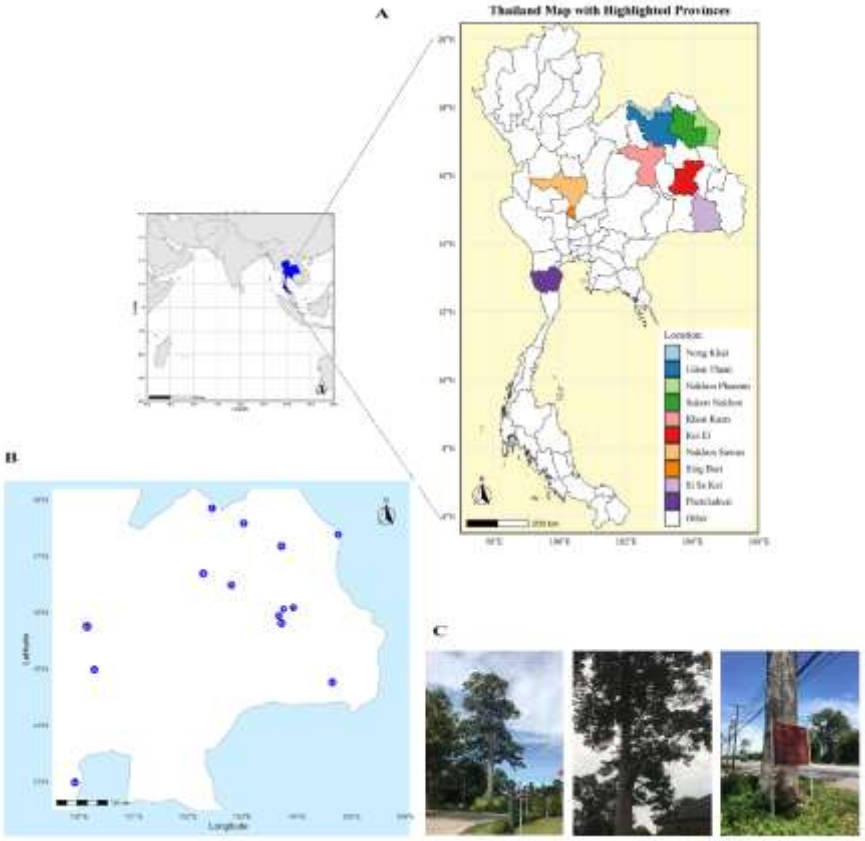

Figure 1. The overview of the sampling sites and morphology of *D. alatus* across Thailand. (A) The native distribution areas of *D. alatus* in Thailand, with highlighted provinces indicating sampling locations. (B) Sampled sites of *D. alatus* in Thailand. (C) The morphology of mature *D. alatus* trees found along roadsides and in public areas of Phetchaburi (Pcb) Province.

Sample tissues were immediately dried in silica gel upon collection to preserve the integrity of the DNA. The dried samples were subsequently ground into fine powder and stored at –20 °C before DNA extraction. Approximately 0.10 g of tissue powder was used for DNA extraction, following the protocol described by Wangsomnuk et al. (2014). DNA pellets were resuspended in 100 µL of TE buffer, and the DNA concentration was quantified using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, USA).

PCR Amplification with SCoT Markers

PCR amplification of *D. alatus* genomic DNA was conducted using 36 SCoT primers, as described by Collard and Mackill (2009), and synthesized by Bio Basic Inc. The efficiency of each primer was initially assessed using DNA samples from six randomly selected individuals originating from six different provinces. Genomic DNA was diluted to a working concentration of 50 ng/µL before amplification. SCoT fingerprinting was conducted according to the protocol described by Mo and Wangsomnuk (2021).

Data Scoring and Statistical Analysis

SCoT banding profiles were scored as binary data, with "1" indicating the presence of a band (dominant allele) and "0" its absence (recessive allele). As SCoT markers are dominant, this binary matrix formed the basis for all subsequent analyses. The complete binary matrix of SCoT loci used for these analyses is

provided in Supplementary Table S1. Marker efficiency was assessed based on several key parameters, including the total number of bands, number and percentage of polymorphic bands (PPL), polymorphic information content (PIC; Serrote et al., 2020) $PIC=1-(p^2+q^2)$, where p and q denote the frequencies of band presence and absence in the binary dataset, resolving power (Rp; Prevost and Wilkinson, 1999), and marker index (MI; Powell et al., 1996). Microsoft Excel and R version 4.2.1 were used to conduct the analyses (R Core Team, 2024), supported by frequently used data manipulation packages, for instance, *tidyverse* and *dplyr*. POPGENE version 1.32 (Yeh et al., 1999) was employed to evaluate genetic diversity. Rudimentary parameters that included the number of observed and effective alleles, Nei's gene diversity, Shannon's information index, and the percentage of polymorphic loci were determined. GenAlEx version 6.5 (Peakall and Smouse, 2012) was employed to evaluate genetic differentiation and molecular variance among populations. A Mantel test was conducted to assess the correlation between genetic and geographic distance. Spatial patterns were evaluated using dbMEMs implemented in the *adespatial* package (Dray et al., 2010). Geographic coordinates were used to build a spatial weighting matrix from the minimum spanning tree. Positive dbMEM eigenfunctions, representing multiscale spatial structure (Dray et al., 2006), were retained and included as predictors in the RDA. Forward selection (9,999 permutations; adjusted R^2) identified significant spatial variables (Blanchet et al., 2008). The global dbMEM model and individual eigenfunctions were tested using 9,999 permutations.

Because SCoT markers follow a dominant-marker model in which band presence reflects the dominant state, population differentiation was quantified using the Φ_PT (PhiPT) statistic derived from the Analysis of Molecular Variance (AMOVA). The significance of Φ_PT was assessed using 999 permutations. Jaccard's coefficient (Jaccard, 1908) was used to calculate pairwise genetic dissimilarities via the *vegan* package in R. The resulting matrix was used to construct a Neighbor-Joining (NJ) dendrogram in MEGA X (Kumar et al., 2018), illustrating the genetic relationships among individuals.

STRUCTURE 2.3.4 (Pritchard et al., 2000), a Bayesian clustering program, was used under the admixture model with correlated allele frequencies, and dominant SCoT data were encoded with the recessive-allele option. We explored K values from 1 to 15, running 40 replicates per K with a burn-in of 10 000 steps followed by 10 000 MCMC iterations. The most likely K was determined using the ΔK method (Evanno et al., 2005) in StructureSelector (Li and Liu, 2018). To check run stability, STRUCTURE was rerun for the selected K with an additional 40 replicates, which were aligned in CLUMPAK using the greedy algorithm to obtain averaged membership coefficients (Q). Individuals with $Q \geq 0.60$ were assigned to a cluster; a stricter cut-off ($Q \geq 0.70$) was also evaluated. Ancestry coefficients were summarized and visualized in R with the packages *dplyr*, *tidyr*, and *ggplot2*.

Pairwise Nei's genetic distances among populations were performed using GenAlEx (version 6.5) to support the clustering results. To assess the congruence between the resulting clusters and those inferred from the STRUCTURE analysis, a Neighbor-Joining (NJ) tree was generated using MEGA X.

RESULTS

Characteristics of SCoT Markers

DNA fingerprinting analysis was conducted using seven selected SCoT primers based on their clarity and reproducibility. Representative gel electrophoresis profiles of 16 individuals from the Sakon Nakhon (SN) population, generated using the SCoT19 and SCoT23 primers, revealed multiple polymorphic bands ranging in size from approximately 280 to 2,300 base pairs (Figure 2). These results indicate that the selected primers were effective in detecting inter-individual genetic variation within a single population.

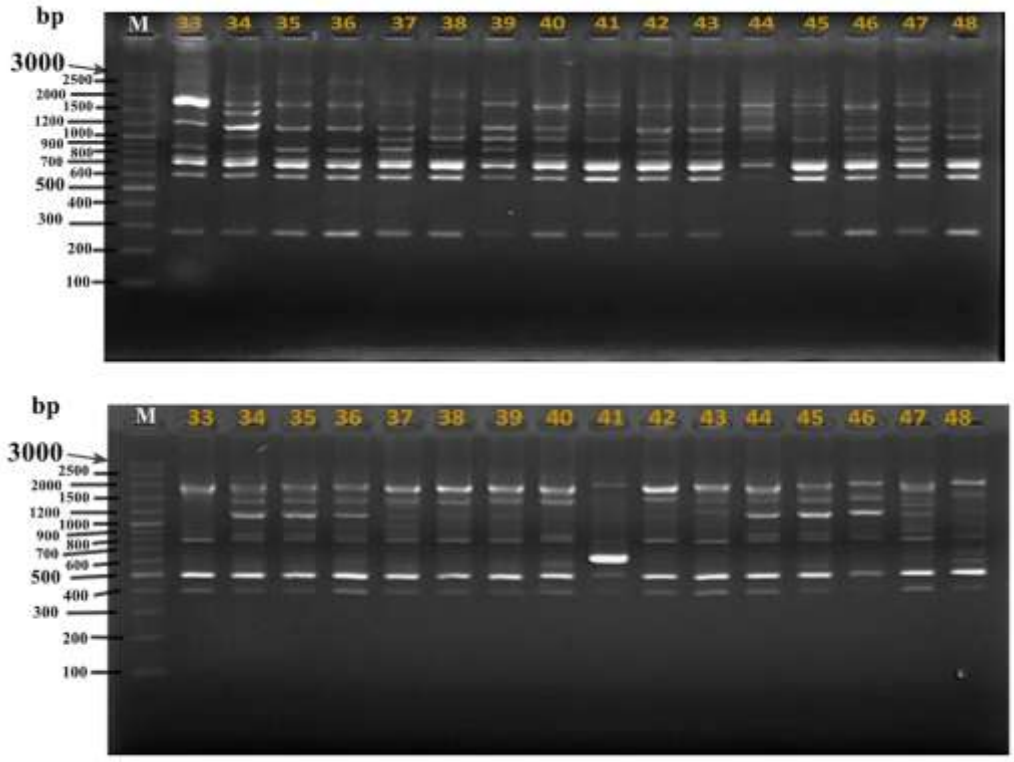

Figure 2. Representative gel electrophoresis profiles of 16 individuals from the Sakon Nakhon (SN) population amplified with primers SCoT19 (top) and SCoT23 (bottom). M, DNA size marker (VC 100 bp Plus DNA Ladder; Vivantis). Numbers 33–48 indicate the sampled individuals from the SN population.

Across all primers, a total of 179 distinct bands were generated, all of which exhibited polymorphism (100% polymorphic rate). The number of amplified bands per primer ranged from 19 to 31. Polymorphic information content (PIC) values ranged from 0.39 to 0.44, with a mean of 0.41. Resolving power (Rp) values ranged between 12.5 and 17.8 (mean = 15.27), while the marker index (MI) varied from 7.815 to 12.03 (mean = 10.27). These results demonstrate that the selected SCoT primers were effective in detecting genetic variation among individuals and populations (**Table 2).**

Table 2. Primer information and marker efficiency parameters of seven selected SCoT primers.

Primer	Primer Sequence (5′–3′)	TNB	PPL	PIC	Rp	MI
SCoT11	AAGCAATGGCTACCACCA	19	100	0.41	12.5	7.815
SCoT19	ACCATGGCTACCACCGGC	28	100	0.40	15.9	11.06
SCoT20	ACCATGGCTACCACCGCG	31	100	0.39	17.8	12.03
SCoT23	CACCATGGCTACCACCAG	22	100	0.44	15.1	9.619
SCoT25	ACCATGGCTACCACCGGG	25	100	0.41	15.5	10.22
SCoT26	ACCATGGCTACCACCGTC	26	100	0.42	16.4	10.85
SCoT32	CCATGGCTACCACCGCAC	28	100	0.40	17.0	11.32

Note: TNB, total number of bands; PPL, percentage of polymorphic bands; PIC, polymorphism information content; Rp, resolving power; MI, marker index.

Genetic diversity at the population level

Genetic diversity indices derived from the SCoT markers showed clear variation among the 14 populations of *D. alatus* (Table 3). The number of observed alleles (Na) ranged from 1.486 to 2.000, with the highest value recorded in the Pcb population and the lowest in the SB population. The effective allele numbers (A_e) varied from 1.263 in population SB to 1.505 in population KK2, suggesting the presence of non-random mating or unequal allele frequencies, where the A_e values were lower than the Na values in all populations. Nei's gene diversity (h) ranged from 0.157 in SB to 0.312 in KK2. At the same time, Shannon's information index (I) spanned from 0.239 in SB to 0.478 in KK2 - the overall mean value of 0.371 points to a moderate level of genetic diversity within populations. The percentage of polymorphic loci (PPL) ranged from 48.60% in SB to 100.00% in Pcb. Other populations exhibiting high PPL values were KK2 (99.44%), SN (97.21%), and NP (96.65%). In contrast, the lowest PPL values were observed in SB (48.60%) and UT (53.07%).

Table 3. Genetic diversity analysis of 14 populations of *D. alatus* based on ScoT markers.

Population	N	Na	A_e	h	I	PPL(%)
NK	7	1.704	1.359	0.219	0.337	70.39
UT	4	1.531	1.308	0.184	0.277	53.07
NP	28	1.967	1.421	0.261	0.407	96.65
SN	31	1.972	1.439	0.274	0.426	97.21
KK1	13	1.827	1.384	0.237	0.369	82.68
KK2	30	1.994	1.505	0.312	0.478	99.44
RE1	20	1.961	1.453	0.285	0.442	96.09
RE2	7	1.687	1.393	0.233	0.352	68.72
RE3	6	1.62	1.348	0.205	0.311	62.01
RE4	14	1.894	1.413	0.253	0.394	89.39
NS	7	1.648	1.347	0.205	0.313	64.80
SB	6	1.486	1.263	0.157	0.239	48.60
Ssk	19	1.933	1.401	0.253	0.397	93.30
Pcb	27	2.000	1.427	0.283	0.449	100.00
Mean		1.801	1.390	0.240	0.371	80.17
SE		0.107	0.089	0.046	0.064	4.90

Note: N, number of individuals in each population; *Na*, observed number of alleles; A_e, effective number of alleles; *h*, Nei's gene diversity; *I,* Shannon's information index; PPL, percentage of polymorphic loci.

Pairwise Nei's unbiased genetic distances ranged from 0.022 to 0.1988. The closest genetic relationship was found between Pcb and KK2 (GD = 0.022), whereas the greatest genetic divergence was observed between SB and RE3 (GD = 0.198) (Table 4).

The spatial analyses revealed a clear spatial signal. The Mantel test showed a moderate but significant positive correlation between genetic and geographic distances (r = 0.262, p = 0.0001), and the Mantel correlogram indicated positive autocorrelation at short distances and negative correlations at larger spatial classes. The dbMEM–RDA identified four significant spatial eigenfunctions (MEM1–MEM4) that together accounted for approximately 5% of the genetic variation, supporting the presence of spatial structuring in *D. alatus* Analysis of Molecular Variance (AMOVA) revealed that 84% of the total genetic variation was attributed to differences within populations, while 16% was attributed to differences among populations (P = 0.001; Table 5). The overall Φ_PT value of 0.156 further indicates moderate genetic differentiation across the sampled populations.

Table 4. Pairwise Population Matrix of Nei Unbiased Genetic Distance

	NK	UT	NP	SN	KK1	KK2	RE1	RE2	RE3	RE4	NS	SB	Ssk	Pcb
NK	0.000													
UT	0.080	0.000												
NP	0.071	0.095	0.000											
SN	0.070	0.076	0.033	0.000										
KK1	0.083	0.078	0.060	0.041	0.000									
KK2	0.083	0.098	0.032	0.041	0.061	0.000								
RE1	0.085	0.103	0.030	0.044	0.069	0.026	0.000							
RE2	0.135	0.145	0.098	0.096	0.120	0.080	0.076	0.000						
RE3	0.147	0.142	0.108	0.082	0.120	0.094	0.097	0.160	0.000					
RE4	0.114	0.109	0.066	0.036	0.081	0.053	0.048	0.105	0.059	0.000				
NS	0.112	0.120	0.072	0.076	0.081	0.077	0.092	0.133	0.165	0.111	0.000			
SB	0.111	0.123	0.115	0.094	0.094	0.114	0.120	0.158	0.198	0.143	0.062	0.000		
Ssk	0.065	0.094	0.036	0.038	0.049	0.038	0.036	0.052	0.105	0.070	0.075	0.097	0.000	
Pcb	0.089	0.087	0.038	0.037	0.068	0.022	0.030	0.073	0.097	0.043	0.102	0.136	0.040	0.000

Table 5. Molecular Variance (AMOVA) analysis of 14 populations of *D. alatus* based on SCoT markers.

Source of Variation	df	SS	MS	Var	%
Among populations	13	1549.574	119.198	5.783	16%
Within populations	205	6434.727	31.389	31.389	84%
Total	218	7984.301	—	37.172	100%

Note: df, degrees of freedom; SS, sum of squares; MS, mean square; Var, estimated variance component; %, percentage of genetic variation.

Population structure and genetic relationships among accessions

Results of Bayesian clustering analysis indicated that the most probable number of genetic clusters (K) among the 219 *D. alatus* individuals was five (K = 5), as determined by a distinct peak in ΔK at K = 5 (ΔK ≈ 240) according to the Evanno method (Figure 3A–B), with detailed LnP(K) and ΔK statistics provided in Supplementary Table S2. When applying a membership coefficient threshold of Q ≥ 0.60 for cluster assignment, 165 individuals (75.34%) were confidently assigned to distinct clusters, whereas 54 individuals

(24.66%) displayed admixed ancestry (Table 6). Among the identified clusters, Cluster 5 contained the highest number of assigned individuals (n = 49; 22.37%), followed by Cluster 1 (n = 42; 19.18%).
The composition of genetic clusters was varied among different populations. Four populations—NK, UT, NS, and SB—showed no evidence of admixture, with all individuals assigned to one or more major clusters (Table 6). NK and UT had individuals distributed between two distinct major clusters, whereas in NS and SB, all individuals belonged to a single cluster, reflecting complete genetic uniformity. By contrast, NP, SN, and Ssk showed high admixture levels, with 50.0%, 48.39%, and 47.37% of individuals, respectively, having Q-values below 0.60. In these three populations, nearly half of the individuals displayed mixed genetic ancestry rather than exclusive membership in a single cluster. Cluster 1 was predominant in NK, UT, KK1, NS, and SB; Cluster 3 dominated in SN, RE3, and RE4; and Cluster 5 was the major group in KK2 and Pcb (Table 6; Figure 3B).

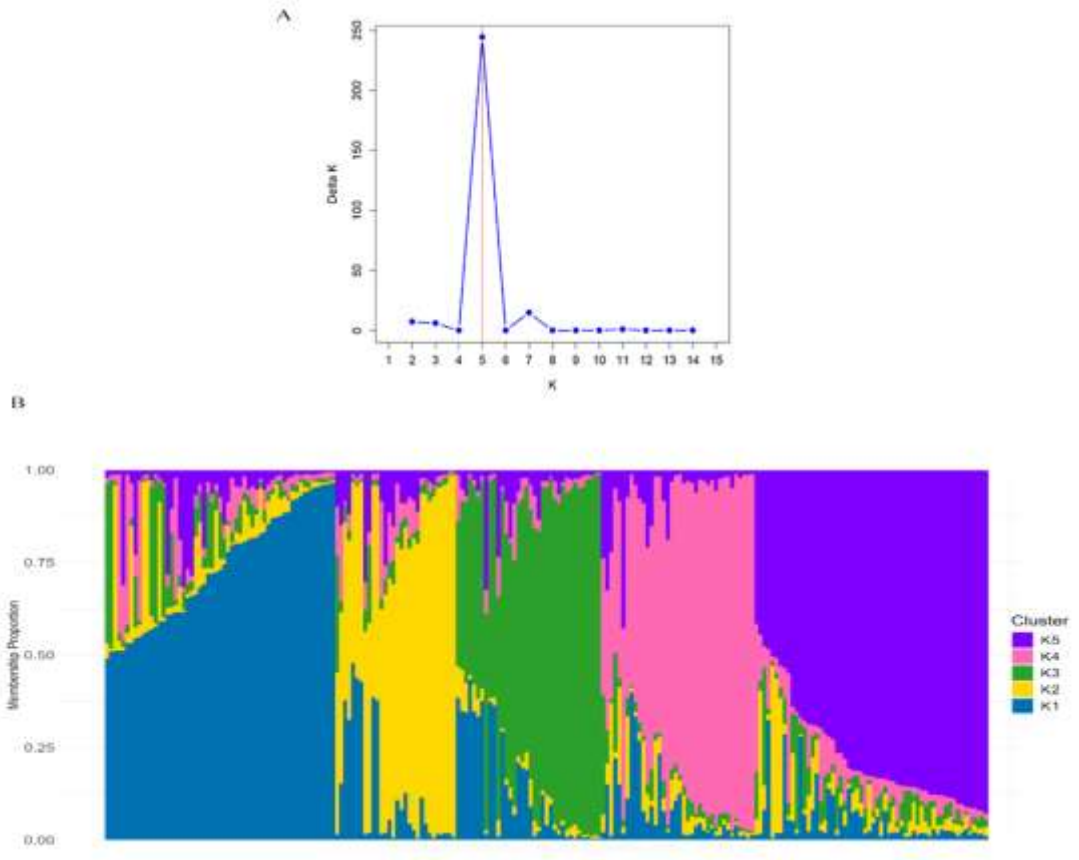

Figure 3. Genetic structure analysis of individuals from 14 *D. alatus* populations.

(A) Estimation of the most likely K value based on the ΔK method (Evanno et al., 2005).
(B) Population structure at K = 5, with CLUMPAK-aligned membership coefficients. Individuals are grouped by dominant cluster and sorted by Q values. Each vertical bar represents one individual, and the colors indicate the proportion of inferred ancestry from each genetic cluster.

Table 6. Individual proportions within different clusters derived from 14 *D.alatus* populations.

Population	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Admixed (n)	Total (n)	Major Cluster	Major Cluster (%)	Admixed (%)
NK	6	0	0	0	1	0	7	2	100.00	0.00
UT	3	0	0	0	1	0	4	2	100.00	0.00
NP	1	11	0	0	2	14	28	3	50.00	50.00
SN	6	0	9	0	1	15	31	3	51.61	48.39
KK1	11	0	0	0	0	2	13	1	84.62	15.38
KK2	0	0	0	11	17	2	30	2	93.33	6.67
RE1	0	8	0	5	0	7	20	2	65.00	35.00
RE2	0	0	0	6	0	1	7	1	85.71	14.29
RE3	0	0	5	0	0	1	6	1	83.33	16.67
RE4	0	0	12	0	1	1	14	2	92.86	7.14
NS	7	0	0	0	0	0	7	1	100.00	0.00
SB	6	0	0	0	0	0	6	1	100.00	0.00
Ssk	2	0	0	7	1	9	19	3	52.63	47.37
Pcb	0	0	0	0	25	2	27	1	92.59	7.41

Notes: Values indicate the number of individuals assigned to each cluster ($Q \geq 0.60$), admixed individuals ($Q < 0.60$), and the corresponding percentages. The major cluster represents the group with the highest number of assigned individuals per population.

To further explore genetic relationships among populations, a Neighbor-Joining (NJ) tree was constructed based on Nei's genetic distance (Figure 4). Populations that shared predominant clusters as defined by STRUCTURE analysis generally clustered together in the tree. For example, NK, UT, KK1, NS, and SB (mainly belonging to Cluster 1) formed a closely related clade. The relative proportion of each genetic cluster within the populations, represented as pie charts adjacent to each terminal node, was largely consistent with the results obtained from the STRUCTURE analysis.

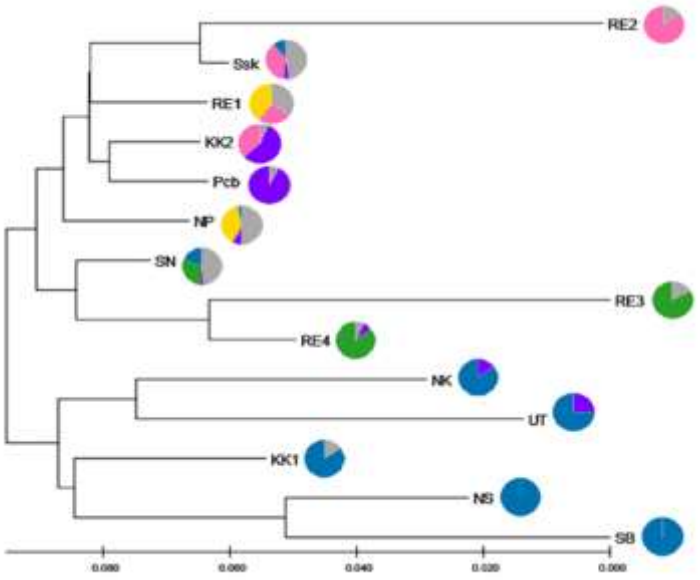

Figure 4 Neighbor-Joining (NJ) tree based on Nei's genetic distance among populations.

Pie charts indicate the proportion of individuals in each STRUCTURE-defined cluster within different populations. Colors representing genetic clusters are**: K1 = blue, K2 = yellow, K3 = green, K4 = pink, K5 = purple, and admixed individuals = gray.**

A distinct NJ tree was constructed based on pairwise Jaccard dissimilarity among individuals, which offered further insights into both intra- and inter-population relationships (Figure 5A). Admixed individuals, particularly those from the NP, RE1, SN, and Ssk populations, were distributed across multiple clades, reflecting genetically heterogeneous backgrounds. When individual tree tips were labeled according to their population of origin (Figure 5B), the majority clustered with members of their respective populations, although inter-population clustering was also evident, especially among populations exhibiting high levels of admixture. This observation suggested the occurrence of historical gene flow and shared ancestral origins among specific populations.

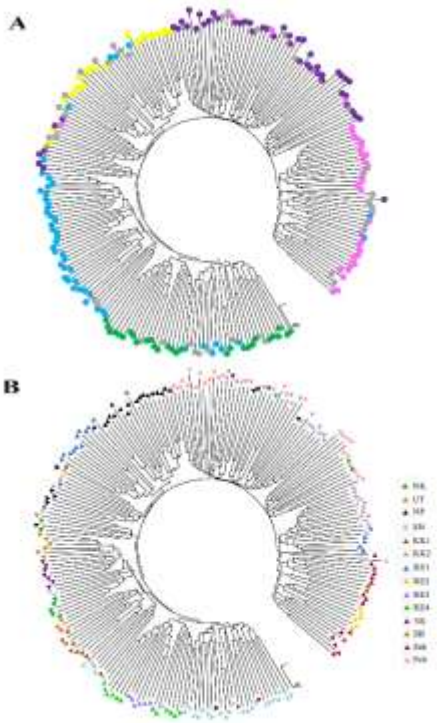

Figure 5: Circular Neighbor-Joining (NJ) trees constructed with 219 individuals of *D. alatus.* (**A**) Constructed NJ tree with entire individuals sorted by cluster groups. Colors representing genetic clusters are**: K1 = blue, K2 = yellow, K3 = green, K4 = pink, K5 = purple, and admixed individuals = gray.**
(**B**) Constructed NJ tree with entire individuals sorted by origin population. Different colors of triangles represent different populations.

DISCUSSION

Effectiveness of SCoT Markers

SCoT markers have demonstrated consistent effectiveness in assessing genetic diversity across various plant taxa, particularly in species lacking extensive genomic resources. Their dominant nature and reliance on the conserved start codon region allow for reproducible and polymorphic amplification. For example, in

Annamocarya sinensis, PIC values ranged from 0.438 to 0.500, with high resolving power (Rp = 9.197), supporting their utility in distinguishing genotypes within populations (Pan et al., 2024).

Despite their limitation in detecting heterozygosity, SCoT markers remain cost-effective and suitable for preliminary genetic assessments. In the present study, all seven SCoT primers produced polymorphic bands in *D. alatus*, with an average PIC of 0.41, which is comparable to that of other tropical trees. The resolving power (Rp) ranged from 12.5 to 17.8, with SCoT20 exhibiting the highest Rp (17.8), indicating a strong capacity for genotype discrimination. SCoT23 recorded the highest PIC value (0.44), underscoring its usefulness in distinguishing between genotypes and further supporting the reliability of these markers in forest genetic studies. These findings align with earlier reports and reinforce the applicability of SCoT markers in conservation and diversity studies of forest tree species.

Genetic Diversity at the Population Level

The Φ_PT estimate points to a moderate level of differentiation among the 14 district-level populations of *D. alatus*. Such a pattern is common in long-lived, predominantly outcrossing tropical tree species, where some local divergence can occur without disrupting overall genetic connectivity. Within Wright's qualitative framework, the degree of differentiation observed here falls squarely in the moderate range and is comparable to that reported for other dipterocarps, including *Shorea ovalis* and *S. leprosula* (Ng et al., 2004). Although the populations show detectable structure, the Φ_PT value indicates that they remain genetically connected rather than strongly isolated. Pollen and seed movement—likely aided by wind or animal dispersal—appears sufficient to sustain gene flow across the landscape. This interpretation is consistent with the AMOVA results, which revealed that most genetic variation occurs within populations, supporting the idea that *D. alatus* maintains a substantial share of its genetic diversity at the local scale.

A significant isolation-by-distance pattern was observed, indicating that geographic distance contributes to genetic differentiation in *D. alatus*. This pattern points to spatially limited gene flow, which aligns with the restricted dispersal range of its winged fruits and the heterogeneous landscape across its natural distribution. Notably, some dipterocarp species do not exhibit isolation-by-distance—for example, *Hopea hainanensis* (Wang et al., 2020)—suggesting that dispersal capacity and landscape influences may vary among closely related taxa. Although gene flow is present among populations, it does not fully mask the spatial genetic structure. A similar pattern has been reported in *Shorea leprosula* (Ng et al., 2004), where moderate gene flow occurs alongside distance-related genetic differentiation. These observations suggest that landscape connectivity and ecological conditions have supported the persistence of *D. alatus* populations, while spatial processes continue to influence their genetic structure.

Pcb and KK2 populations, for example, presented a considerable level of genetic diversity, as indicated by their elevated proportions of polymorphic loci. Such patterns may result from comparatively large population sizes or greater genetic adaptation to diverse local environments. The Pcb population originates from a natural habitat that once supported a large number of *D. alatus* trees, although its size has since declined substantially. The diversity observed in Pcb likely reflects the legacy of historical genetic variation maintained over generations, even as recent population reductions may threaten its long-term persistence. In contrast, KK2 was established by Khon Kaen University as part of a targeted ex situ conservation initiative aimed at safeguarding the species' genetic resources. KK2 exhibited the highest diversity among all populations studied (PPL = 99.44%, h = 0.312, I = 0.478), a pattern more likely linked to human-mediated introductions from multiple, though undocumented, sources. Although the precise origins are unknown, assembling individuals from genetically distinct backgrounds within the same managed site has likely enhanced allelic richness and heterozygosity, contributing to the elevated diversity indices observed in this population. On the other hand, populations such as SB and UT, which have comparatively low proportions of polymorphic loci, may have experienced recent genetic bottlenecks or habitat fragmentation, both of which are known to reduce allelic richness gradually.

The results from the AMOVA showed that the populations contained 84% of the total genetic variation, a tendency commonly observed in long-lived and predominantly outcrossing tropical tree species. Similar patterns were observed in *Shorea leprosula* (70.2%) and *S. parvifolia* (66.2%) (Cao et al., 2006), as well as in fragmented populations of *Hopea hainanensis* (77.4%; Wang et al., 2020). More comprehensive surveys, for example, those by Oluwajuwon et al. (2022), reported that intrapopulation variation reaches up to 89%, thereby strengthening the concept that tropical tree populations conserve a significant portion of the species' genetic diversity. This internal variation is likely a crucial factor in sustaining resilience and adaptability, even when confronting environmental change.

Populations like Pcb and KK2 could serve as important reservoirs of genetic diversity and may deserve attention in future conservation planning. In contrast, populations with lower diversity—especially those in fragmented or isolated habitats—may be more vulnerable to genetic erosion. Conservation strategies should therefore aim to safeguard both genetically rich and at-risk populations, with an emphasis on maintaining landscape connectivity and supporting natural regeneration processes.

While SCoT markers are limited in their ability to distinguish co-dominant alleles or provide fine-scale resolution, they offer several advantages when working with non-model species that lack genomic references. Their capacity to detect polymorphisms without prior sequence data makes them especially useful for preliminary assessments of genetic variation in tropical forest trees. However, results based on dominant markers should be interpreted carefully, particularly when concluding population structure or levels of heterozygosity.

Population Structure and Admixture

Bayesian clustering analysis using STRUCTURE identified five genetic clusters (K = 5) among the 219 individuals sampled. Approximately 75.34% were predominantly assigned to a single cluster ($Q \geq 0.60$), while the remaining 24.66% showed evidence of admixture. This mix of clearly defined clusters and admixed individuals aligns with findings in *D. turbinatus* along environmental gradients (Duc et al., 2023) and in *Shorea leprosula*, where forest fragmentation followed by reforestation resulted in secondary contact zones (Ng et al., 2004).

Some populations—SB and NS—were entirely assigned to a single cluster, suggesting long-term isolation or limited gene exchange, possibly due to founder effects. These genetically distinct populations may represent Evolutionarily Significant Units (ESUs) with unique adaptive potential. In contrast, NP, SN, and Ssk showed high levels of admixture, with nearly half of the individuals having Q-values below 0.60. Such patterns may reflect ongoing gene flow, anthropogenic disturbances, or secondary contact between divergent lineages.

It was noteworthy that individuals identified as admixed by STRUCTURE did not uniformly align with those exhibiting high average Jaccard dissimilarity (Supplementary Table S3). This inconsistency underscored the fundamental differences between model-based and distance-based analytical approaches. Whereas STRUCTURE determined cluster membership based on allele frequency distributions, Jaccard dissimilarity assessed genetic divergence through the presence or absence of genetic bands. The integration of these complementary methodologies thereby provided a more comprehensive perspective on the genetic structure of *D. alatus*. A sensitivity check using a stricter assignment threshold ($Q \geq 0.70$) naturally reduced the number of individuals that could be confidently assigned to clusters. In three populations (UT, NP, and SN), this stricter cutoff also led to a smaller dominant cluster or fewer detectable clusters. These changes, however, were confined to individual-level assignments. The broader clustering pattern—particularly the identity of the major cluster in each population and the relationships among populations—remained

essentially unchanged. This stability indicates that the population structure inferred for *D. alatus* is robust to the choice of Q threshold (Supplementary Table S4).

Spatial Genetic Patterns

The Neighbor-Joining (NJ) tree based on Nei's genetic distance was largely congruent with the clusters inferred from STRUCTURE, indicating that allele frequency variation provides a robust signal of population-level relationships in *D. alatus*. Populations sharing a dominant ancestry, such as NK, UT, KK1, NS, and SB, consistently clustered together, reflecting their genetic homogeneity. In contrast, the Jaccard-based NJ tree, which relies on binary presence/absence data, revealed finer-scale heterogeneity, particularly within populations such as NP, RE1, SN, and Ssk. The dispersal of admixed individuals across multiple clades in this tree suggests ongoing or historical gene flow, possibly facilitated by human-mediated movement of propagules or overlapping flowering periods between divergent lineages.

These contrasting patterns highlight the complementary value of the two distant metrics. Nei's genetic distance is more effective for detecting broad-scale structure shaped by historical processes, whereas Jaccard dissimilarity is sensitive to recent or individual-level variation. Considering both perspectives provides a more nuanced understanding of genetic relationships in *D. alatus*, supporting conservation strategies that address both long-term lineage divergence and the maintenance of within-population diversity.

Conservation implications

There are significant conservation implications from the observed genetic structure of *D. alatus* populations. In populations such as UT, NK, SB, and NS, having fairly uniform genetic profiles may indicate localized adaptation and provide stable baselines for tracking future genetic changes. In situ conservation should be highlighted in these locations to protect their potentially region-specific traits. Conversely, elevated allelic diversity was demonstrated in admixed populations, such as SN, Ssk, and NP, which may enhance their ability to manage environmental pressures. These populations possess noteworthy potential for restoration seed sourcing, where genetic resilience is paramount. Broadhurst et al. (2008) noted that when planning effective conservation, genetic distinctiveness and diversity within populations must be embraced.

Many *D. alatus* populations are located in disjointed, human-generated landscapes, such as village forests, temple grounds, and roadside remnants (Thianchai et al., 2013). These landscapes have both ecological and social impacts, highlighting the species' susceptibility and underscoring their importance in grassroots conservation. Embracing local knowledge and combining it with genetic evidence will potentially reinforce the effects of long-term preservation and advance pragmatic, community-grounded strategies for maintaining this environmentally and financially essential species.

CONCLUSION

This is the first study to utilize elevated allelic diversity and SCoT markers to investigate genetic variation in *D. alatus*. The results indicate moderate genetic differentiation, high within-population diversity, and a complex isolation-by-distance pattern, suggesting that gene flow continues across the species. The genetic structure appears to be shaped by both natural connectivity and human activities, with a combination of distinct and admixed populations. Notably, many samples were drawn from community-managed landscapes such as village forests, temple groves, and schoolyards, highlighting the role of local stewardship in genetic conservation beyond formal protected areas. While dominant markers like SCoT have their limitations, they have proven effective in capturing meaningful patterns in this non-model species. The approach provides a practical tool to inform both in situ conservation and community-based management, particularly as environmental pressures on tropical tree species continue to intensify.

DATA AVAILABILITY.
The datasets generated and analyzed during this study—including raw gel images, coarse GPS coordinates, and analysis scripts—are available from the corresponding author upon reasonable request.

ACKNOWLEDGMENTS
We sincerely thank the journal reviewers for their thoughtful and constructive comments, which greatly improved the clarity and quality of this manuscript. This research was supported by Khon Kaen University and the Plant Genetic Conservation Project under the Royal Initiative of Her Royal Highness Princess Maha Chakri Sirindhorn. We are also deeply grateful to the local community members, landowners, and officials who generously provided guidance, access, and assistance during fieldwork and sample collection across multiple regions of Thailand.

CONFLICTS OF INTEREST
The authors declare that they have no conflict of interest.

REFERENCES

Blanchet FG, Legendre P, Borcard D (2008). Forward selection of explanatory variables. *Ecology*. 89: 2623–2632. doi: 10.1890/07-0986.1

Broadhurst LM, Lowe A, Coates DJ, Cunningham SA, et al. (2008). Seed supply for broadscale restoration: maximizing evolutionary potential. *Evol Appl*. 1: 587-597. doi: 10.1111/j.1752-4571.2008.00045.x

Caneva G, Bartoli F, Zappitelli I, Savo V (2020). Street trees in Italian cities: story, biodiversity and integration within the urban environment. *Rend Lincei Sci Fis Nat*. 31: 411-417. doi: 10.1007/s12210-020-00907-9

Cao CP, Finkeldey R, Siregar I, Siregar U, et al. (2006). Genetic diversity within and among populations of *Shorea leprosula* Miq. and *Shorea parvifolia* Dyer (Dipterocarpaceae) in Indonesia detected by AFLPs. *Tree Genet Genomes*. 2: 225-239. doi: 10.1007/s11295-006-0046-0

Changtragoon S, Boontawee B (1999). The study of genetic diversity of *Dipterocarpus alatus* by isoenzyme gene markers, pp. 107-14. In: Seminars on *Dipterocarpus alatus* and Dipterocarpaceae, 17-18 Nov 1999, Kasetsart Univ, Bangkok, Thailand.

Changtragoon S, Ongprasert P, Tangmitcharoen S, Diloksampan S, et al. (2013). Country report on forest genetic resources of Thailand. In: The State of the World's Forest Genetic Resources. FAO, Rome.

Chokthaweepanich H, Butdavapee S, Phongkaew P (2022). Genetic diversity and relationships of *Dipterocarpus alatus* Roxb. (Dipterocarpaceae) on island in freshwater lake and nearby mainland. *Agric Nat Resour*. 56: 351-362.

Collard BCY, Mackill DJ (2009). Start codon targeted (SCoT) polymorphism: a simple, novel DNA marker technique for generating gene-targeted markers in plants. *Plant Mol Biol Rep*. 27: 86-93. doi: 10.1007/s11105-008-0060-5

Dray S, Legendre P, Peres-Neto PR (2006). Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecol Model*. 196: 483-493. **doi:** 10.1016/j.ecolmodel.2006.02.015

Dray S, Bauman D, Blanchet G, Borcard D, et al. (2010). adespatial: Multivariate multiscale spatial analysis. R package version 0.3-28. Available at [https://cran.r-project.org/package=adespatial] Accessed October 10, 2025.

Duc NM, Hoang NH, Giang TTH, Huong NTT, et al. (2023). Genetic variation and evolutionary history of the threatened *Dipterocarpus turbinatus* C.F.Gaertn. detected using microsatellites. *Diversity*. 15: 894. doi: 10.3390/d15080894

Dyrmose AMH, Turreira-García N, Theilade I, Meilby H (2017). Economic importance of oleoresin (*Dipterocarpus alatus*) to forest-adjacent households in Cambodia. Nat Hist Bull Siam Soc. 62: 67-84.

Evanno G, Regnaut S, Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol.* 14: 2611-2620. doi: 10.1111/j.1365-294X.2005.02553.x

Jaccard P (1908). Nouvelles recherches sur la distribution florale. *Bull Soc Vaudoise Sci Nat*. 44: 223-270.

Kamyo T, Asanok L (2019). Modeling habitat suitability of *Dipterocarpus alatus* (Dipterocarpaceae) using MaxEnt along the Chao Phraya River in Central Thailand. *For Sci Technol*. 16: 1-7. doi: 10.1080/21580103.2019.1687108

Kumar S, Stecher G, Li M, Knyaz C, et al. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 35: 1547-1549. doi: 10.1093/molbev/msy096

Li YL, Liu JX (2018). StructureSelector: a web-based software to select and visualize the optimal number of clusters using multiple methods. *Mol Ecol Resour*. 18: 176-177. doi: 10.1111/1755-0998.12719

Ly V, Nanthavong K, Pooma R, Barstow M, et al. (2023). *Dipterocarpus alatus* (amended version of 2017 assessment). IUCN Red List of Threatened Species, e.T33007A231194805. Available at: https://www.iucnredlist.org/species/33007/231194805. Accessed Aug 9, 2025.

Mo XC, Wangsomnuk PP (2021). A modified non-liquid nitrogen protocol for extraction of high-quality genomic DNA from the inner bark tissues of *Dalbergia cochinchinensis* (Fabaceae). *Genet Mol Res*. 20: GMR18836. doi: 10.4238/gmr18836

Ng KK, Lee SL, Koh CL (2004). Spatial structure and genetic diversity of two tropical tree species with contrasting breeding systems and different ploidy levels. *Mol Ecol*. 13: 657-669.

Nguyen THH, Tran NN, Nguyen TLH (2012). Study of the genetic diversity of *Dipterocarpus alatus* Roxb. by using RAPD markers. *Viet J For Sci*. 3.

Oluwajuwon TV, Attafuah R, Offiah CJ, Krabel D (2022). Genetic variation in tropical tree species and plantations: a review. *Open J For*. 12: 207-230. doi: 10.4236/ojf.2022.123019

Orwa C, Mutua A, Kindt R, Jamnadass R, et al. (2009). Agroforestree Database: a tree reference and selection guide version 4.0. World Agroforestry Centre (ICRAF), Kenya.

Pan H, Deng L, Zhu K, Shi D, et al. (2024). Evaluation of genetic diversity and population structure of *Annamocarya sinensis* using SCoT markers. *PLoS One*. 19: e0309283. doi: 10.1371/journal.pone.0309283

Peakall R, Smouse PE (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*. 28: 2537-2539.

Powell W, Morgante M, Andre C, Hanafer M, et al. (1996). The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Mol Breed*, 2: 225-238. doi: 10.1007/BF00564200.

Prevost A, Wilkinson MJ (1999). A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. *Theor Appl Genet*. 98: 107-112.

Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics*. 155: 945-959. doi: 10.1093/genetics/155.2.945

R Core Team (2024). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rai MK (2023). Start codon targeted (SCoT) polymorphism marker in plant genome analysis: current status and prospects. *Planta*. 257: 34. doi: 10.1007/s00425-023-04067-6

Serrote CML, Reiniger LRS, Silva KB, Rabaio SMD, et al. (2020). Determining the polymorphism information content of a molecular marker. *Gene*, 726: 144175. doi: 10.1016/j.gene.2019.144175

Tam NM, Duy VD, Due NM, Giap VD, et al. (2014). Genetic variation in and spatial structure of natural populations of *Dipterocarpus alatus* (Dipterocarpaceae) determined using single sequence repeat markers. *Genet. Mol. Res*., 13: 5378-5386. doi: 10.4238/2014.July.24.17.

Thianchai U, Diloksumpun S, Tangmitrcharoen S (2013). Current status and conservation strategies of the genetic resources of *Dipterocarpus alatus* outside its natural habitat by the Royal Forest Department. *Thai J For Sci*. 32(Suppl): 119-32. (in Thai)

Vu DD, Bui TTX, Nguyen MD, Shah SNM, et al. (2019). Genetic diversity and conservation of two threatened dipterocarps (Dipterocarpaceae) in southeast Vietnam. *J For Res.*, 30: 1823-1831. doi: 10.1007/s11676-018-0735-1

Wangsomnuk PP, Ritthitum W, Ruttawat B, Wangsomnuk P (2014). Comparative analysis of DNA extracted from mature leaves of rubber tree and application for seventeen tropical plant species for PCR amplification. *AAB Bioflux*. 6: 44-56.

Wati M, Puspitasari N, Hairah U, Widians JA, et al. (2024). Dipterocarpaceae trunk texture classification using two-stage convolutional neural network-based transfer learning model. *Int J Electr Comput Eng*. 14: 6874-6882. doi: 10.11591/ijece.v14i6.pp6874-6882

Wang C, Ma X, Ren M, Tang L (2020). Genetic diversity and population structure in the endangered tree *Hopea hainanensis* (Dipterocarpaceae) on Hainan Island, China. *PLoS One*. 15: e0241452. doi: 10.1371/journal.pone.0241452

Yeh FC, Yang RC, Boyle T (1999). POPGENE Version 1.32: Microsoft Windows-based freeware for population genetic analysis. Univ Alberta, Canada.

SUPPLEMENTARY MATERIAL

Table S1. Binary presence–absence matrix of SCoT loci used for genetic analyses. Column headers indicate the specific SCoT loci (primer–fragment combinations)

Table S2. Summary of LnP(K) statistics, ΔK, and Evanno method results (K = 1–15).

Table S3. Sample codes with their corresponding average Jaccard genetic distances (JD) and cluster assignments (Clu/Adx) inferred from STRUCTURE analysis at K = 5. Individuals were grouped based on the highest membership probability, considering only those with $Q \geq 0.60$. Individuals showing $Q < 0.60$ were classified as admixed (Adx).

Table S4. Sensitivity analysis of individual cluster assignments using a stricter membership threshold ($Q \geq 0.70$), compared with the primary threshold of $Q \geq 0.60$.